The Anatomy of the Human Genome

VICTOR A. McKUSICK, M.D.
Baltimore, Maryland

More than 350 genes have been assigned to specific chromosomes. These include more than 110 assigned to the X chromosome, more than 240 assigned to specific autosomes, and at least one assigned to the Y chromosome. (Even man's 25th chromosome, that of the mitochondrion, is being mapped.) Almost all the assignments to specific autosomes were made in the last decade. About half of these were made by study of clones derived from interspecific (e.g., man-mouse) somatic cell hybrids. Over a fifth were made by family linkage studies. Chromosomes 1 and 6 are rather extensively mapped. The genes of over 40 autosomal disorders have been specifically localized. The comparative anatomy, functional anatomy, developmental anatomy and even applied anatomy of the human genome is becoming better known.

In the last 10 years one of the most rapid and significant developments in our understanding of the genetics of man has been the mapping of the human chromosomes [1]. Ten years ago about 80 loci had been assigned to the X-chromosome—by the observation of the characteristic X-linked pedigree patterns of distinctive human traits. However, only three genes had been assigned to specific autosomes: the Duffy blood group locus to chromosome 1 by family linkage studies of a variant form (heteromorphism) of chromosome 1; the alpha-haptoglobin locus to chromosome 16 by family studies of familial translocations involving that chromosome and of a heteromorphism of chromosome 16; the thymidine kinase locus to chromosome 17 by somatic cell hybridization. By June 1976 at least one gene had been assigned to each of man's 22 autosomes; today [2] about 250 specific gene loci have been assigned to specific autosomes (**Figure 1**). The mapping of genes on chromosome 1 and on chromosome 6 is extensive, and a small segment of the short arm of chromosome 11 has been mapped in elegant detail (see later). Great intellectual satisfaction can be derived from the fact that we now know that the Rh blood group locus is on the distal part of the short arm of chromosome 1; the ABO blood group locus is near the end of the long arm of chromosome 9; the serum albumin locus is on the long arm of chromosome 4; the locus of the sickle cell gene is on the short arm of chromosome 11, and so on. The chromosomal localization of the genes coding for insulin, growth hormone and several other substances of clinical interest are in the process of being mapped [19].

About one-fourth of the autosomal assignments were achieved by family studies [3, 4], i.e., studies of linkage between marker traits and a marker chromosome, or studies of linkage between two marker traits, one of which had already been assigned to a specific autosome. (**Figure**

From the Department of Medicine, Johns Hopkins Hospital and Johns Hopkins University School of Medicine, Baltimore, Maryland 21205. Requests for reprints should be addressed to Dr. Victor A. McKusick, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205. Manuscript accepted March 10, 1980.

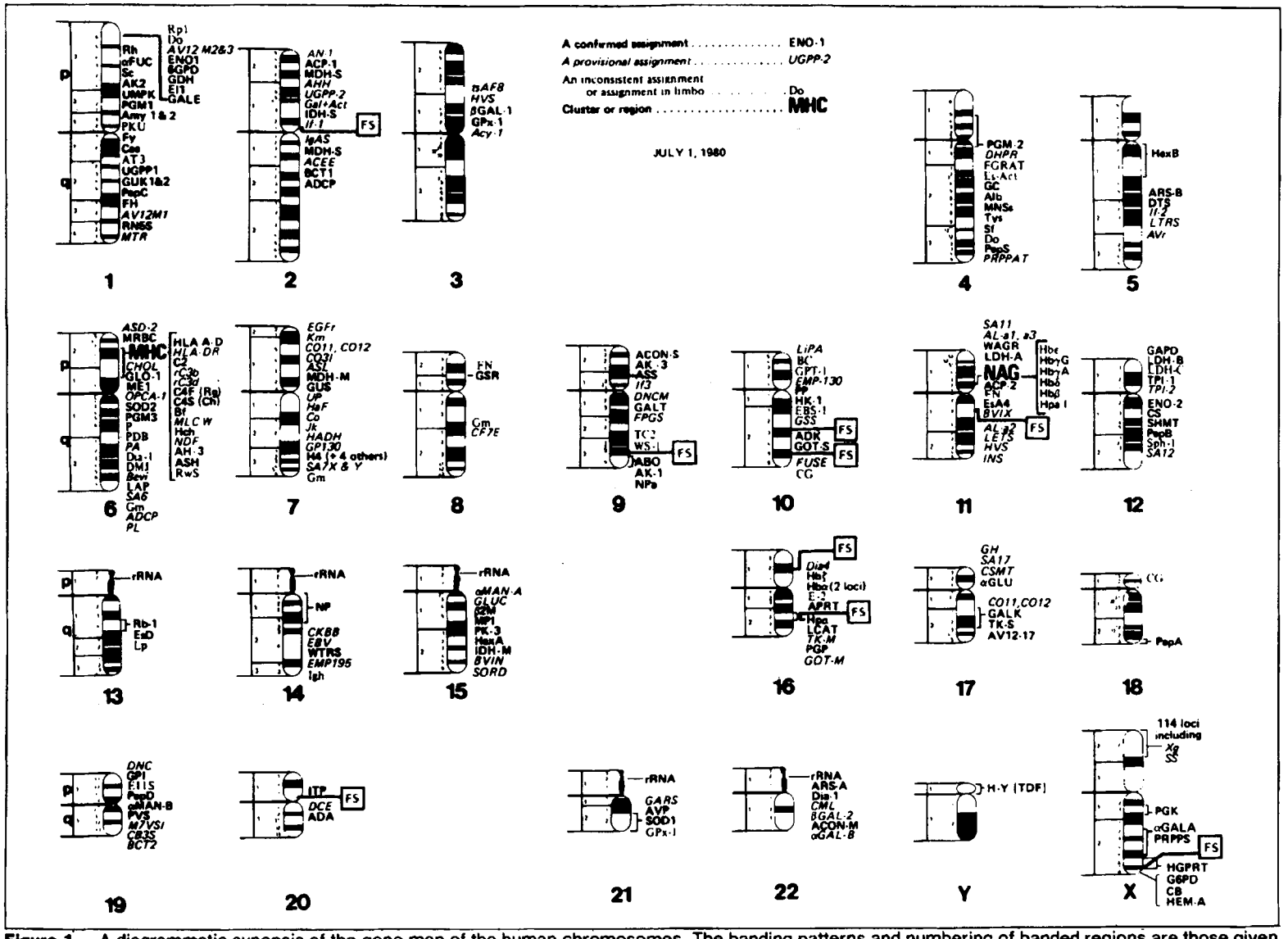

Figure 1. A diagrammatic synopsis of the gene map of the human chromosomes. The banding patterns and numbering of banded regions are those given in the International System for Human Cytogenetics Nomenclature 1978. A assignment is considered confirmed if found in two laboratories or several families; it is considered provisional if based on evidence from only one laboratory. Inconsistent assignments based on conflicting evidence and assignments for which the evidence is weaker than that for provisional assignment are separately indicated (also termed "tentative" or "in limbo"). See Key for gene locus symbols.

KEY TO GENE SYMBOLS SHOWN IN FIGURE 1 (with indication of chromosomal and in some instances regional location of the gene: e.g., 9q34 means band 34 of the long arm of chromosome 9, 2p23 means band 23 of chromosome 2)

ABO = ABO blood group—9q34
ACEE = Acetylcholinesterase expression—chr. 2
ACON-M = Aconitase, mitochondrial—22q11-22qter
ACON-S = Aconitase, soluble—9pter-9p13
ACP1 = Acid phosphatase-1—2p23
ACP2 = Acid phosphatase-2—11p12-cen
ACY1 = Aminoacylase-1—3pter-q13
ADA = Adenosine deaminase—20q13-qter
ADCP1 = Adenosine deaminase complexing protein-1—chr. 6
ADCP2 = Adenosine deaminase complexing protein-2—chr. 2
ADK = Adenosine kinase—10q11-q24
AH3 = Adrenal hyperplasia III (21-hydroxylase deficiency) (6p2105-6p23)
AHH = Aryl hydrocarbon hydroxylase—2p
AK1 = Adenylate kinase-1 (soluble)—9q34
AK2 = Adenylate kinase-2 (mitochondrial)—1p32-1p34
AK3 = Adenylate kinase-3 (mitochondrial)—9pter-p13
AL = Lethal antigen: 3 loci—a1,a3 on 11p13-pter; a2 on 11q13-qter
Alb = Albumin—4q11-q13
AMY1 = Amylase, salivary—1p1
AMY2 = Amylase, pancreatic—1p1
An1 = Aniridia, type 1 (chr.2; linked to ACP1)
APRT = Adenine phosphoribosyltransferase—16q
ARS-A = Arylsulfatase A—chr. 22
ARS-B = Arylsulfatase B—chr. 5
ASD2 = Atrial septal defect, secundum type (chr. 6; linked to HLA)
ASH = Asymmetric septal hypertrophy (chr. 6; linked to HLA)
ASL = Argininosuccinate lyase—7pter-q22
ASS = Argininosuccinate synthetase—chr. 9
AT3 = Antithrombin III (chr. 1)
AV12M1 = Adenovirus-12 chromosome modification site-1—1q42-43
AV12M2 = Adenovirus-12 chromosome modification site-2—1p36
AV12M3 = Adenovirus-12 chromosome modification site-3—1q21
AV12-17 = Adenovirus-12 chromosome modification site-17—17q21-q22
AVP = Antiviral protein—21q21-qter
AVr = Antiviral state regulator—chr. 5
β2M (B2M) = Beta-2-microglobulin—15q14-q21
BCT-1 = Branched chain amino acid transferase-1—chr. 12
BCT-2 = Branched chain amino acid transferase-2—chr. 19
BEVI = Baboon M7 virus infection—chr. 6
BF = Properdin factor B—chr. 6 (in MHC)
BVIN = BALB virus induction, N-tropic—chr. 15
BVIX = BALB virus induction, xenotropic—chr. 11
C2 = Complement component-2—chr. 6 (in MHC)
C4F = Complement component-4 fast—chr. 6 (in MHC)
C4S = Complement component-4 slow—chr. 6 (in MHC)
Cae = Cataract, zonular pulverulent (chr. 1; linked to Fy)
CB = Colorblindness (deutan and protan) (Xq26-Xqter)
CB3S = Coxsackie B3 virus susceptibility—chr. 19
CF7E = Clotting factor VII expression (chr. 8)
CG = Chorionic gonadotropin (chr. 10 and 18; chr. 5 or 6)
Ch = Chido blood group—same as C4S
CHOL = Hereditary hypercholesterolemia—chr. 6 (?linked to HLA)
CKBB = Creatine kinase, brain type—chr. 14
CML = Chronic myeloid leukemia—22q12
Co = Colton blood group (chr. 7)
CO11 = Collagen I alpha-1 chain—chr. 7 and 17
CO12 = Collagen I alpha-2 chain—chr. 7 and 17
CO31 = Collagen III alpha-1 chain—chr. 7
CS = Citrate synthase, mitochondrial—chr. 12
CSMT = Chorionic somatomammotropin—(chr. 17)
DCE = Desmosterol-to-cholesterol enzyme—chr. 20
DHPR = Quinoid dihydropteridine reductase—chr. 4
Dia-1 = NADH-diaphorase—chr. 22
DIA-4 = Diaphorase-4—chr. 16
DMJ = Juvenile diabetes mellitus (chr. 6; ?linked to HLA)
DNC = Lysosomal DNA-ase—chr. 19
DNCM = Cytoplasmic membrane DNA—9qh
Do = Dombrock blood group (?chr. 1 or 4)
DTS = Diphtheria toxin sensitivity—5q15-5qter
E2 = Pseudocholinesterase-2—16cen-q22
E11S = Echo 11 sensitivity—19q
EBS1 = Epidermolysis bullosa, Ogna type (chr. 10)
EBV = Epstein-Barr virus integration site—chr. 14
EGFR = Epidermal growth factor, receptor for—chr. 7
E11 = Elliptocytosis-1—(1p; linked to Rh)
EMP130 = External membrane protein-130—chr. 10
EMP195 = External membrane protein-195—chr. 14
ENO1 = Enolase-1—1p36-1pter
ENO2 = Enolase-2—chr. 12

Cont'd on page 270

Es-Act = Esterase activator—chr. 4 or 5
EsA4 = Esterase-A4—11cen-q22
EsD = Esterase D—13q14
FGPS = Folylpolyglutamate synthetase—chr. 9
FGRAT = Formylglycinamide ribotide amidotransferase—chr. 4 or 5
FH = Fumarate hydratase—1q42-qter
FN = Fibronectin—chr. 8, 11
FS = Fragile site, observed in cultured cells, with or without folate deficient medium, or BrdU—2q11; 10q23; 10q25; 11q13; 16p124; 16q22; 20p11; Xq27
αFUC (FUCA) = Alpha-L-fucosidase—1p32-p34
FUSE = Polykaryocytosis inducer—chr. 10
Fy = Duffy blood group—1q13
Gal+-Act = Galactose + activator—chr. 2
αGALA = Alpha-galactosidase A (Fabry disease)—Xq22-Xq24
αGALB = Alpha-galactosidase B—22q13-qter
βGAL-1 = Beta-galactosidase-1—3pter-q13
βGAL-2 = Beta-galactosidase-2—22q13-qter
GALE = Galactose-4-epimerase—1p21-1pter
GALK = Galactokinase—17q21-q22
GALT = Galactose-1-phosphate uridyltransferase—9p13 or 9p22
GAPD = Glyceraldehyde-3-phosphate dehydrogenase—12p122-12pter
GARS = Glycinamide ribonucleotide synthetase—chr. 21
GC = Group-specific component—4q11-4q13
GH = Growth hormone—chr. 17
GDH = Glucose dehydrogenase—1p21-1pter (1p32-1pter)
αGLU (GLUA) = Alpha-glucosidase—chr. 17
GLUC = Neutral alpha-glucosidase C—chr. 15
GLO1 = Glyoxylase I—6p21-6p22
GOT-M = Glutamate oxaloacetate transaminase, mitochondrial—chr. 16
GOT-S = Glutamate oxaloacetate transaminase, soluble—10q24-q26
G6PD = Glucose-6-phosphate dehydrogenase—Xq28
GP130 = Granulocyte glycoprotein—7q22-7qter
GPI = Glucosephosphate isomerase—chr. 19
GPT1 = Glutamate pyruvate transaminase, soluble—chr. 10
GPx1 = Glutathione peroxidase-1—3p13-q12
GSR = Glutathione reductase—8p21
GSS = Glutamate-gamma-semialdehyde synthetase—chr. 10
Gm = Immunoglobulin heavy chain—chr. 6,7,8 (see Igh)

KEY TO GENE SYMBOLS (Cont'd)

GUK1 & 2 = Guanylate kinase-1 & 2—1q32-1q42
GUS = Beta-glucuronidase—chr. 7
H4 = Histone H4 and 4 other histone genes—chr. 7
HADH = Hydroxyacyl-CoA dehydrogenase—chr. 7
HaF = Hageman factor—7q35
Hbα(HBA) = Hemoglobin alpha chain—chr. 16
Hbβ(HBB) = Hemoglobin beta chain—11p1205-11p1208
Hbδ(HBD) = Hemoglobin delta chain—11p1205-11p1208
Hbγ(HBG) = Hemoglobin gamma chains—11p1205-11p1208
Hbε(HBE) = Hemoglobin epsilon chain—11p1205-11p1208
Hbζ(HBZ) = Hemoglobin zeta chain—chr. 16
Hch = Hemochromatosis (chr. 6; linked to HLA)
HEM-A = Classic hemophilia—Xq26-Xqter
HexA = Hexosaminidase A—15q22-15qter
HexB = Hexosaminidase B—5cen-q13
HGPRT = Hypoxanthine-guanine phosphoribosyltransferase—Xq26-Xq27
HK1 = Hexokinase-1—10pter-q24
HLA (A-D) = Human leukocyte antigens—6p2105-p23
HLA-DR = Human leukocyte antigen, D-related—6p2105-p23
Hpα = Haptoglobin, alpha—chr. 16
HpaT = Hpa I restriction endonuclease polymorphism—11p1205-11p1208
HVS = Herpes virus sensitivity (chr. 3 and 11)
H-Y = Y histocompatibility antigen (Y chr.)
IDH-M = Isocitrate dehydrogenase, mitochondrial—15q21-15qter
IDH-S = Isocitrate dehydrogenase, soluble—2q11 or 2q32-2qter
If1 = Interferon-1—2p23-qter
If2 = Interferon-2 (chr. 5)
If3 = Interferon-3 (chr. 9)
IgAS = Immunoglobulin heavy chains attachment site—chr. 2
Igh = Immunoglobulin heavy chains (mu, gamma, alpha)—chr. 14 (see Gm)
Ins = Insulin—chr. 11
ITP = Inosine triphosphatase—20p
Jk = Kidd blood group—7q
Km = Kappa immunoglobulin light chains, Inv (chr. 7)
LAP = Laryngeal adductor paralysis—(chr. 6; linked to HLA)
LCAT = Lecithin-cholesterol acyltransferase—(16q22; linked to Hp alpha)
LDH-A = Lactate dehydrogenase A—11p1203-p1208
LDH-B = Lactate dehydrogenase B—12p121-p122
LDH-C = Lactate dehydrogenase C—(12p; linked to LDH-B in pigeon)
LIPA = Lysosomal acid lipase-A—chr. 10
Lp = Lipoprotein—Lp—chr. 13
LTRS = Leucyl-tRNA synthetase—chr. 5
β2M (B2M) = Beta-2-microglobulin—15q22-15qter (15q12-15q21)
M7VS1 = Baboon M7 virus sensitivity-1—chr. 19
αMAN-A = Cytoplasmic alpha-D-mannosidase—15q11-15qter
αMAN-B = Lysosomal alpha-D-mannosidase—19pter-19q13
MDH-M = Malate dehydrogenase, mitochondrial—7p22-q22
MDH-S = Malate dehydrogenase, soluble—2p23
ME1 = Malic enzyme, soluble—6p21-q16
MHC = Major histocompatibility complex—6p2105-6p23
MLC-W = Mixed lymphocyte culture, weak (chr. 6)
MNSs = MNSs blood group—4q
MPT = Mannosephosphate isomerase—15q22-qter
MRBC = Monkey red blood cell receptor—chr. 6
MTR = 5-Methyltetrahydrofolate: L-homocysteine S-methyltransferase, or tetrahydropteroyl-glutamate methyltransferase—chr. 1
NAG = Non-alpha globin region—12p1205-1208
NDF = Neutrophil differentiation factor (chr. 6)
NP = Nucleoside phosphorylase—14q13
NPa = Nail-patella syndrome—(9q3; linked to ABO)
OPCA1 = Olivopontocerebellar atrophy I—(chr. 6; linked to HLA)
P = P blood group (chr. 6)
PA = Plasminogen activator (chr. 6)
PDB = Paget disease of bone—(chr. 6; ?linked to HLA)
PepA = Peptidase A—18q23-18qter
PepB = Peptidase B—12q21
PepC = Peptidase C—1q25, or 1q42
PepD = Peptidase D (chr. 19)
PepS = Peptidase S—4pter-4q12
6PGD = 6-phosphogluconate dehydrogenase—1p34-1pter
PGK = Phosphoglycerate kinase—Xq13
PGM1 = Phosphoglucomutase-1—1p32; 1p221-p311; 1p33-1p34
PGM2 = Phosphoglucomutase-2—4p14-q12
PGM3 = Phosphoglucomutase-3—6q
PGP = Phosphoglycolate phosphatase—16p
PK3 = Pyruvate kinase-3—15q14-qter
PKU = Phenylketonuria (1p; linked to AMY)
PL = Prolactin
PRPPAT = Phosphoribosylpyrophosphate amidotransferase—4pter-q21
PRPPS = Phosphoribosylpyrophosphate synthetase—X chr.
PP = Inorganic pyrophosphatase—10pter-q24
PVS = Polio virus sensitivity—19q
RB1 = Retinoblastoma-1—13q12-q14; 13q21-22
rC3b = Receptor for C3b—chr. 6 (in MHC)
rC3d = Receptor for C3d—chr. 6 (in MHC)
Rg = Rodgers blood group—same as C4F
Rh = Rhesus blood group (1p32-pter)
RN5S = 5S RNA gene(s)—1q42-1q43
RP1 = Retinitis pigmentosa-1 (chr. 1)
rRNA = Ribosomal RNA—13p12, 14p12, 15p12, 21p12, 22p12
RwS = Ragweed sensitivity—(chr. 6; ?linked to HLA)
SA6 = Surface antigen 6—chr. 6
SA7 = Species antigen 7—7p12-pter
SA11 = Surface antigen 11-11p
SA12 = Surface antigen 12—chr. 12
SA17 = Surface antigen 17—chr. 17
SA21 = Surface antigen 21—chr. 21
Sc = Sclanna blood group—(1p32-1p34)
Sf = Stoltzfus blood group—(4q; linked to MNSs)
SHMT = Serine hydroxymethyltransferase—chr. 12
SOD1 = Superoxide dismutase, soluble—21q211
SOD2 = Superoxide dismutase, mitochondrial—6q21
SORD = Sorbitol dehydrogenase—15pter-q21
Sph 1 = Spherocytosis, Denver type (8p11 or chr. 12)
SS = Steroid sulfatase (?Xp22-Xpter)
TC2 = Transcobalamin II—9q (?linked to ABO)
TDF = Testis determining factor—prob. same as H-Y
Tk-M = Thymidine kinase, mitochondrial—chr. 16
TK-S = Thymidine kinase, soluble—17q21-q22
TPI-1 & 2 = Triosephosphate isomerase-1 & 2—TPI-1 on 12p12.2-12pter
tsAF8 = Temperature sensitive (AF8) complementing—chr. 3
Tys = Sclerotylosis—(4q; linked to MNSs)
UGPP1 = Uridyl diphosphate glucose pyrophosphorylase-1—1q21-q23
UGPP2 = Uridyl diphosphate glucose pyrophosphorylase-2—chr. 2
UMPK = Uridine monophosphate kinase—1p32
UP = Uridine phosphorylase—chr. 7
WAGR = Wilms tumor—aniridia/ambiguous genitalia/mental retardation—11p13
WTRS = Tryptophanyl-tRNA synthetase—chr. 14
WS1 = Waardenburg syndrome-1—(chr. 9; ?linked to ABO)
Xg = Xg blood group (X chr.,?Xp2)

2 shows the type of map that can be constructed from family data on linkage.) About 60 percent of the autosomal assignments have been made by somatic cell hybridization. About 6 percent of the assignments have been made independently by both family studies and somatic cell hybridization. Other assignments have been made by various methods such as linkage disequilibrium, nucleic acid hybridization and restriction enzyme mapping.

The largest number of genes have been assigned to the X chromosome; indeed, before 1968 the X chromosome was the only one whose specific genetic content was known at all. This lead resulted from the fact that X localization of genes is deducible from the well-known pedigree pattern of traits such as colorblindness and hemophilia—a fact first pointed out in 1911 by Columbia cytogeneticist E. B. Wilson. Over 110 gene loci are now confidently assigned to the human X chromosome by the family pedigree method.* In the last decade, other methods have permitted mapping of some of these genes to specific regions of the X chromosome. For example, we now know that the genes for glucose-6-phosphate dehydrogenase (G6PD), colorblindness and classic hemophilia are clustered at the distal end of the long arm and that the X-linked blood group Xg^a is situated at the distal end of the short arm.

At least one gene (or family of genes) has been assigned to the Y chromosome. Since 1959 when, from chromosomal studies of normal men and women, and people with the Turner and Klinefelter syndromes, as well as other abnormalities of sex chromosomes, it has been known that the Y chromosome carries a testis-determining factor (TDF); a close correlation was found between the presence or absence of a Y chromosome and the presence or absence of testes. More recently TDF has been identified with H-Y, the Y-determined histocompatibility antigen which is also a differentiation antigen critical to the development of testes from the indifferent gonad.

The cartographic simile is a natural one for the process of assigning particular genes to particular chromosomes, identifying the precise localization of genes on the physical map as defined by the banding pattern of metaphase chromosomes, and measuring in genetic terms (recombination units) the distance between pairs of loci. An equally apt simile is an anatomic one; the chromosomes and the specific genes they carry are aspects of human anatomy. Furthermore, to amplify on the analogy, we now have a morbid anatomy, a comparative anatomy, a functional anatomy, a developmental anatomy and even, to some extent, an applied anatomy of the human genome.

As noted earlier, recent methods for dissecting the human genome include restriction endonuclease mapping of segments of DNA isolated by hybridization with nucleic acid probes provided by recombinant DNA

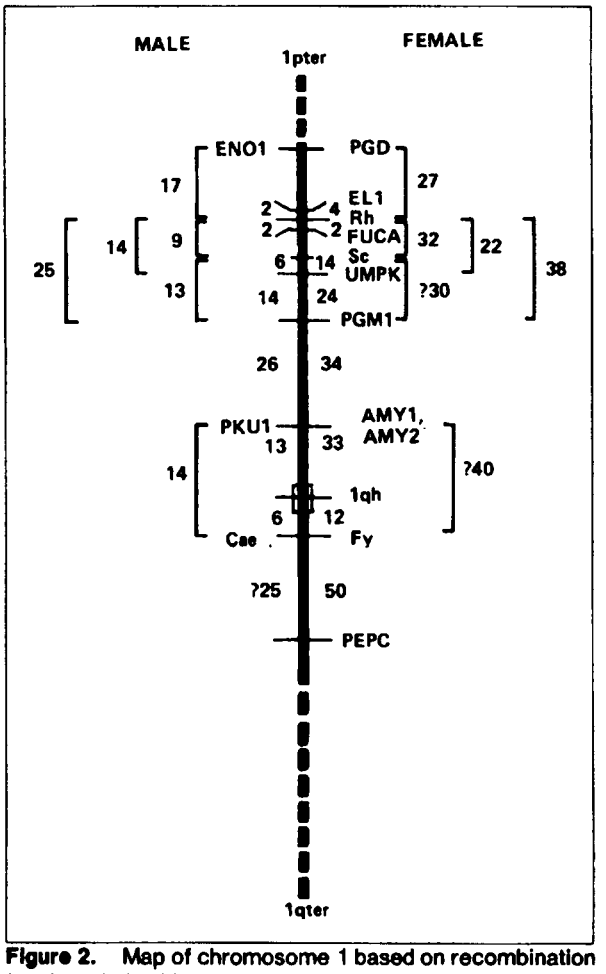

Figure 2. Map of chromosome 1 based on recombination fractions derived from family studies. Recombination fractions are greater in females than in males. See Key for Figure 1, for significance of gene symbols. *pter* and *qter* refer to the ends of the short and long arms, respectively. *Iqh* refers to a heterochromatic region on the long arm adjacent to the centromere. (The information given by a map such as this is of particular usefulness in the application of linkage to prenatal and premorbid diagnosis. The recombination fraction is a statement of probability that traits determined at two given loci will separate in transmission from one generation to the next.) From report of Committee on Chromosome 1 (P.J.L. Cook and John L. Hamerton, Chairmen), Human Gene Mapping Workshop No. 5, Edinburgh, 1979 [21].

* In five successive editions of my computerized "Mendelian Inheritance in Man" [5], all known autosomal and X-linked loci are purportedly listed, with description of mutant phenotypes and peculiarities of the genetics and with key bibliographic references. The fifth edition was published in 1978. The most recent update lists almost 1,400 autosomal and more than 110 X chromosome loci, with an additional 1,450 autosomal loci and 110 X loci which are thus far only provisionally characterized. The total of 3,070 loci represents only a small part of the 50,000 to 100,000 structural genes that man is estimated to have.

technology. These techniques provide a structural delineation of the human genome which, in terms of resolution, is analogous to microscopic anatomy; by comparison, delineation by other techniques is analogous to gross anatomy. The best example of this fine structure analysis in man is that involving the segment of the short arm of chromosome 11 which carries the six non-alpha hemoglobin genes. The fine intragenic structure of the individual globin genes has also been revealed by restriction mapping.

The Mobid Anatomy of the Human Genome. The specific chromosomal localization of the genes mutant in at least 40 autosomally inherited diseases of man is now known (**Table I**). Included are some 10 lysosomal storage diseases, such as Tay-Sachs disease, two disorders of the urea cycle, five disorders of carbohydrate metabolism, several immune deficiency disorders, phenylketonuria in both its classic and atypical forms, and neoplasms such as retinoblastoma and Wilms tumor.

The Functional Anatomy of the Human Genome. It turns out that structural genes for enzymes in sequential steps of particular metabolic pathways are, as a rule, not syntenic* (**Table II**). This nonsynteny holds for the urea cycle (three enzymes mapped), the tricarboxylic acid cycle (five enzymes mapped), galactose metabolism (three enzymes mapped) and the phosphogluconate pathway (three enzymes mapped). A suggestion of clustering occurs in the case of the glycolytic sequence in which four enzymes are found to be determined by genes on chromosome 12. Whereas in bacteria the enzymes for sequential steps in some metabolic pathways are determined by linked genes, thus assuring coordinate activity, the situation is quite different in man. This, perhaps, is not surprising since in eukaryotic cells many enzymes, e.g., glyceraldehyde-3-phosphate dehydrogenase, function in more than one metabolic pathway and coordinate activity might be disruptive.

The Comparative Anatomy and Evolution of the Human Genome. The karyotype of the great apes bears close similarities to each other and to that of man as revealed by banding pattern and by over-all morphology. It, therefore, came as no great surprise that a number of autosomal loci that are syntenic (on the same chromosome) in man are also syntenic (and on the chromosome which is judged to be homologous on morphologic grounds) in the chimpanzee, orangutan, baboon and rhesus. More surprising, perhaps, is the

* James H. Renwick did a valuable semantic service to genetics by introducing *syntenic* (noun = *synteny*) for the situation when two loci are on the same chromosome, whether linked or not. *Linkage*, on the other hand, refers to the situation of two genetic loci sufficiently close that something less than 50 percent recombination occurs between traits determined by genes at the two loci. For example, on the X chromosome the Xg blood group locus and the hemophilia A locus are syntenic but not linked.

TABLE I Genetic Diseases for Which the Mutation Has Been Mapped to a Specific Autosome*

A. Disorders of carbohydrate metabolism	
Fucosidosis	1p
Galactosemia	9p
Galactokinase deficiency	17q
Galactose-4-epimerase deficiency	1p
Glycogen storage disease II	17
B. Disorders of amino acid metabolism	
Classic phenylketonuria	1(?)
Dihydropteridine reductase deficiency	4
C. Diseases of lipid metabolism	
Norum disease (LCAT deficiency)	16q
D. Lysosomal storage diseases	
(Fabry disease	Xq)
Generalized gangliosidosis	3
Glycogen storage disease II	17
Lysosomal acid phosphatase deficiency	11p
Mannosidosis	19
Metachromatic leukodystrophy	22q
MPS II	X
MPS VI	5
MPS VII	7
Sandhoff disease	5q
Tay-Sachs disease	15q
Wolman disease	10
E. Urea cycle disorders	
Argininosuccinicaciduria	7
Citrullinemia	9
(Ornithine transcarbamylase deficiency	X)
F. Congenital nonspherocytic hemolytic anemia	
Glucosephosphate isomerase deficiency	19
Hexokinase deficiency	10
Triosephosphate isomerase deficiency	12p
G. Other hematologic disorders	
Elliptocytosis	1p
Sickle cell anemia	11p
Thalassemias	11p, 16p
H. Immune deficiency diseases	
Adenosine deaminase deficiency	20q
Nucleoside phosphorylase deficiency	14q
C2 deficiency	6p
C4 deficiency	6p
Transcobalamin II deficiency	9q
I. Endocrinopathies	
Congenital adrenal hyperplasia (21-hydroxylase deficiency)	6p
J. Malignant neoplasm	
Wilms tumor (WAGR syndrome)	11p
Retinoblastoma	13q
Chronic myeloid leukemia	22q
K. Miscellaneous	
Analbuminemia	4q
Cataract, zonular pulverulent	1
Hemochromatosis	6p
Nail-patella syndrome	9q
Olivopontocerebellar atrophy I	6
Tetrahydrofolate methyltransferase deficiency	1
Dihydroxyadenine urolithiasis (APRT deficiency)	16

* The numbers refer to chromosome carrying the particular locus, with arm when known. p = short arm; q = long arm. Three X-linked disorders are listed for sake of completeness in certain categories. The list is not exhaustive.

TABLE II The Chromosomal Localization of Structural Genes for Enzymes in Five Pathways

A. Galactose metabolism	
Galactose-4-epimerase	1p
Galactokinase	17q
Galactose-1-phosphate uridyltransferase	9p
B. The urea cycle	
Argininosuccinate synthetase	9
Argininosuccinate lyase	7
Ornithine transcarbamylase	X
C. Tricarboxylic acid cycle	
Aconitase, mitochondrial	22q
Aconitase, soluble	9q
Isocitrate dehydrogenase, mitochondrial	15q
Isocitrate dehydrogenase, soluble	2q
Fumarase	1q
Malate dehydrogenase, mitochondrial	7
Malate dehydrogenase, soluble	2p
Citrate synthase, mitochondrial	12
D. Glycolysis	
Entry to glycolytic sequence	
Hexokinase	10
Phosphoglucomutase-1	1p
Phosphoglucomutase-2	4
Phosphoglucomutase-3	6q
Mannosephosphate isomerase	15q
First stage of glycolysis	
Glucosephosphate isomerase	19
Triosephosphate isomerase 1 & 2	12p
Second stage of glycolysis	
Glyceraldehyde-3-phosphate dehydrogenase	12p
Phosphoglycerate kinase	Xq
Enolase-1	1p
Enolase-2	12
Lactate dehydrogenase-A	11p
Lactate dehydrogenase-B	12p
(Lactate dehydrogenase-C	12p)
E. Phosphogluconate pathway (pentose phosphate pathway; hexose monophosphate shunt)	
Glucose-6-phosphate dehydrogenase	Xq
6-Phosphogluconate dehydrogenase	1p
Glyceraldehyde-3-phosphate dehydrogenase	12p

NOTE: p = short arm; q = long arm.

TABLE III Paralogous Genes in Man Presumably Derived by Regional Duplication

Loci	Chromosome
Ag lipoproteins (five loci)	?
Amylases, pancreatic and salivary	1
Complement components-2, -4F, and -4S and properdin factor B	6
Complement component-6 and -7	?
Globin, alpha (2 loci)	16
Globins, nonalpha (6 loci)	11p
Growth hormone; chorionic somatomammotropin	17
Guanylate kinase-1 and -2	1q
Histones (5 loci)	7q
Immunoglobulin heavy chains	?
Major histocompatibility complex	6p
Parotid salivary proteins (6 or probably 8 loci)	?

degree of homology of synteny discovered between man and his more remote relative, the mouse [6].

The ultimate in homology is observed in the case of the X chromosome of mammals. Ohno's law of the evolutionary conservatism of the mammalian X chromosome in mammals has no known exception; the gene content of the X seems to be identical in all mammals, placental and marsupial. This implies that no fundamental evolutionary change has occurred in the mammalian X chromosome in nearly 100 million years. Presumably, lyonization, developing as a mechanism of compensation for discrepancy in the number of X chromosomes in males and females ("dosage compensation") "froze" the genetic constitution of the X chromosome in the form it had millions of years ago when the mechanism originated.

One can speak of two types of homologous genes [6]: Orthologous genes (or loci) are those which exist in different species but arose from a common ancestral gene. Paralogous genes (or loci) are those which occur in the same species, i.e., is in one person, having been derived from a common ancestral gene by the process of regional (tandem) duplication, translocation or tetraploidization. About a dozen of the presently mapped genes (**Table III**) are apparently paralogous and probably arose by tandem duplication inasmuch as they are closely situated on the same chromosome and the gene products are similar.

Other possibly paralogous genes are situated on different chromosomes; examples are listed in **Tables IV** and **V**. As shown in Table 5, the list includes seven enzymes that have both soluble (cytoplasmic or cytosolic) and mitochondrial forms, determined by unlinked genes on separate chromosomes. These loci may be nonsyntenic paralogous loci that derived from a common ancestral gene by an ancient process of tetraploidization or perhaps by tandem duplication followed by translocation. The two forms of thymidine kinase (on chromosomes 16 and 17) may be examples. Some may, however, have had quite different origins and may then have evolved to serve the same or similar function. For example, the mitochondrial form of superoxide dismutase is similar to that of bacteria, but except for its similar catalytic function the soluble form is widely different in all characteristics, including amino acid sequence. Yet other enzymes that have both soluble and mitochondrial forms (fumarase hydratase is thus far the only clear examples in man) are determined by a single structural gene [7] and the electrophoretic difference is the consequence of post-translational processing.

(DNA is not limited to the nucleus; mitochondria,

TABLE IV Some Possibly Paralogous Genes in Man Originating by Translocation or Tetraploidization

Acid phosphatase-1 (red cells)	2p
Acid phosphatase-2 (lysosomal)	11p
Branched chain amino acid transferase-1	2
Branched chain amino acid transferase-2	19
Enolase-1	1p
Enolase-2	12
Esterase-A4 } Homology not established	11q
Esterase-D } Homology not established	13q
β-Galactosidase-1	3
β-Galactosidase-2	22
α-globin of hemoglobin	16
β-globin of hemoglobin	11p
Lactate dehydrogenase-A	11p
Lactate dehydrogenase-B	12p
α-D-mannosidase, cytoplasmic	15q
α-D-mannosidase, lysosomal	2q
Peptidase A } Homology not established	18q
Peptidase B } Homology not established	12q
Peptidase C } Homology not established	1q
Peptidase D } Homology not established	19
Peptidase S } Homology not established	4
Phosphoglucomutase-1	1p
Phosphoglucomutase-2	4
Phosphoglucomutase-3	6
Uridyl diphosphate glucose pyrophosphorylase-1	1
Uridyl diphosphate glucose pyrophosphorylase-2	2

NOTE: p = short arm; q = long arm.

TABLE V Some Possibly Paralogous Genes in Man

A. Mitochondrial and cytosolic (soluble) isozymes determined by nonsyntenic loci	
Aconitase, mitochondrial	22q
Aconitase, soluble	9q
Adenylate kinase-1 (soluble)	9pter-9p13
Adenylate kinase-2 (mitochondrial)	1p
Adenylate kinase-3 (mitochondrial)	9q34
Glutamate oxaloacetate transaminase, mitochondrial	6
Glutamate oxaloacetate transaminase, soluble	10q
Isocitrate dehydrogenase, mitochondrial	15q
Isocitrate dehydrogenase, soluble	2q
Malate dehydrogenase, mitochondrial	7
Malate dehydrogenase, soluble	2p
Malic enzyme, mitochondrial	(not 6)
Malic enzyme, soluble	6
Superoxide dismutase, mitochondrial	6q
Superoxide dismutase, soluble	21q
Thymidine kinase, mitochondrial	16
Thymidine kinase, soluble	17
B. Mitochondrial and cytosolic (soluble) isozymes determined by a single locus with post-translational processing	
Fumarate hydratase (fumarase), mitochondrial and soluble	1q

NOTE: p = short arm; q = long arm.
• Although AK-1 and AK-3 are strictly speaking syntenic, they are on separate arms of chromosome 9.

which are thought to have originated from an ancient bacterium that took up symbiotic intracellular residence, have a single circular chromosome (cf. Figure 3, [8]) and have the machinery for translation and protein synthesis (9). All the mitochondrial enzymes shown in Table V are coded by nuclear genes (nDNA), are synthesized in the cytoplasm and are then transported into mitochondria. The human mitochondrial chromosome has been mapped by restriction endonuclease techniques similar to those used in bacteria (cf. Figure 1 [10]). Mitochondrial DNA (mtDNA) codes for certain components of the cytochrome oxidase system of mitochondria. At least one mutation in human mtDNA [11]—that responsible for chloramphenicol resistance—has been identified in cultured cells [12]. On the whole, the mitochondrion has a relatively limited genetic repertoire and most of its structural and functional elements are coded by nDNA.)

Hemoglobin is an example of a polymeric nonenzymic protein comprised of subunits coded by genes on different chromosomes (chromosomes 11 and 16). Lactate dehydrogenase is an example of a polymeric enzyme that is constituted by units coded by genes on different chromosomes (chromosome 11 for LDH A; chromosome 12 for LDH B). Many such "multilocus" enzymes show the phenomenon of greater or even predominant activity of one locus in a particular tissue whereas in another tissue the other locus has predominant activity. Since the two forms of the enzyme have somewhat different functional properties, this permits specialization appropriate to the given tissue. As pointed out by Harris [13], regulatory mechanisms specific to each tissue must have evolved pari passu with the evolution of the separate structural loci.

The Developmental Anatomy of the Human Genome. Although genes responsible for successive enzymes in metabolic pathways are, as a rule, not linked, the anatomic arrangement of some genes on the chromosomes seems to have ontogenetic significance: Genes which are sequentially activated during development are linked. This is most clearly illustrated in the case of the non-alpha globin cluster. The embryonic, fetal and adult globin genes exist in duplicate, in each case, possibly representing a fail-safe system. The linear, 5′ to 3′ arrangement of the six genes along a short segment of the short arm of chromosome 11 are epsilon-2,

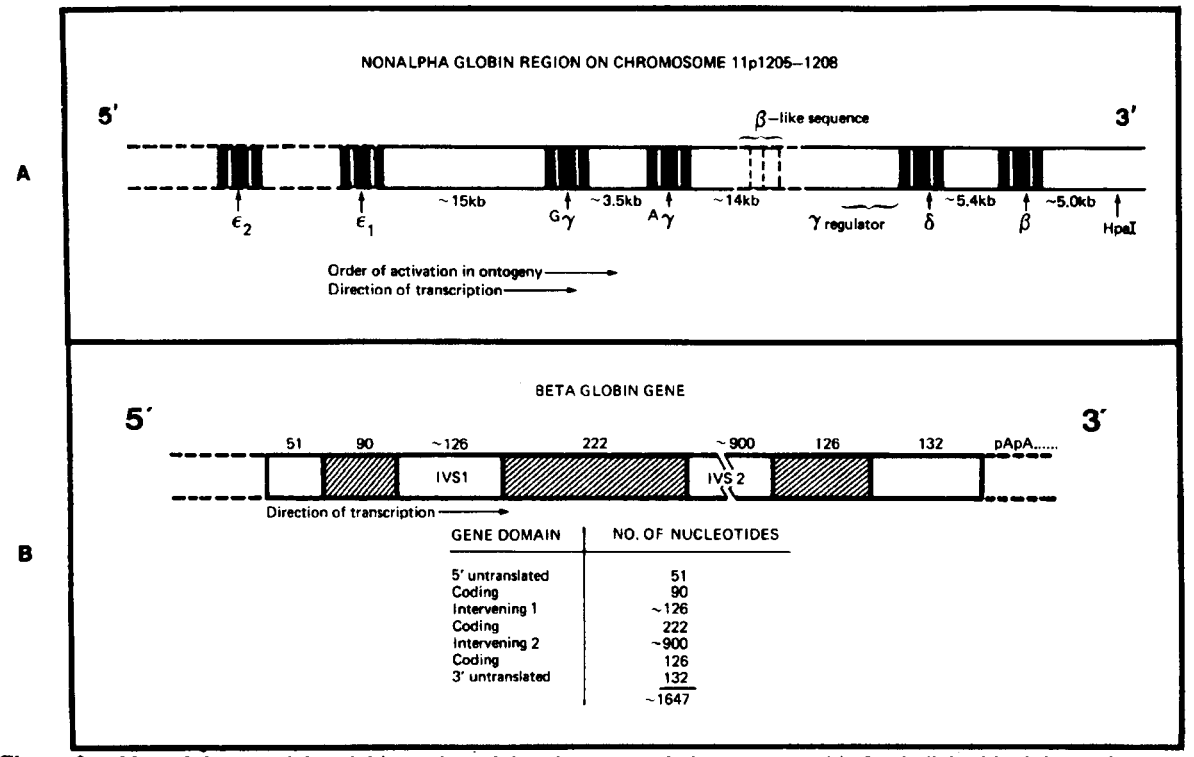

Figure 3. Map of the nonalpha-globin region of the short arm of chromosome 11. **A**, six linked loci determine nonalpha-polypeptides of hemoglobin: two epsilon loci for embryonic hemoglobins ($\alpha_2\epsilon_2$), two loci for fetal homoglobins ($\alpha_2\gamma_2$) and two loci, δ and β, for adult hemoglobins A ($\alpha_2\beta_2$) and A_2 ($\alpha_2\delta_2$). Study of various deletions suggests that a segment between $^A\gamma$ and δ exercises a regulatory function in $^\gamma$polypeptide synthesis. The gene labelled ϵ_2 differs considerably from ϵ_1. It was designated $\psi\beta 2$ by Fritsch et al. [20], who referred to the β-like sequence between the γ genes and the δ gene as $\psi\beta 1$. HpaI is a nucleic acid polymorphism in a noncoding region about 5kb from the 3' end of the gene. Each of the genes has three coding regions separated by two noncoding intervening sequences, as shown for the β gene in B. After Forget et al. [18] with additions by personal communication from Dr. Tom Marniatis, California Institute of Technology, and others. **B**, the coding and noncoding segments of the beta-globin gene. The 5' end of the gene corresponds to the NH2-end of the polypeptides and is transcribed first. It is designated 5' because it corresponds to the 5' end of messenger RNA.

epsilon-1, G-gamma, A-gamma, delta and beta (**Figure 3A**). The embryonic (epsilon) genes are active in the embryo, the fetal (gamma) genes in the fetus and the adult (delta and beta) genes after birth. Furthermore, the two gamma genes are present in the proportion of about 7:3 throughout the fetal period but, when the switch to adult genes occurs, in the postnatal period the small amount of fetal hemoglobin detectable in the normal adult is predominantly of the A-gamma type in a G:A proportion of 2:3 [14]. Of course, the beta globin gene is much more active than the delta gene in adult life, again illustrating the progressive switching of activity from the 5' to the 3' end of the nonalpha globin cluster during ontogeny. (The same restriction endonuclease techniques that have allowed mapping of the cluster of nonalpha globin (beta-like) genes have also revealed their internal structure as shown in **Figure 3B.**)

Differentiation of lymphocytes, in relation to the immunoglobulin loci, is anatomically unique. Seemingly, somatic recombination occurs between the gene(s) for the constant portion and those for the variable portion of the immunoglobulin heavy chain [15]. This permits the generation of a great diversity of antibody genes. Allelic exclusion—the activity of only one of the two homologous chromosomes—occurs in the process.

The Applied Anatomy of the Human Genome. Inevitably, detailed knowledge of the anatomy of the human genome will find extensive application during this last vintade* of the 20th century. Already, it has found usefulness in connection with prenatal diagnosis of the linkage principle. At least four conditions—myotonic dystrophy, classic hemophilia, the 21-hydroxylase deficiency type of congenital adrenal hyperplasia and sickle cell anemia—all not directly determinable in cultured amniotic cells, can now be diagnosed in the fetus by linkage with, respectively, secretor, G6PD, HLA and HpaI restriction fragment

* I use the word, doubtless a neologism, to mean for 20 years what decade means for 10.

length polymorphism. Each of these four markers shows polymorphism that can be detected in amniotic fluid or cells. The HpaI polymorphism [16] represents a variant in the DNA itself resulting in differences in cleavage by site-specific restriction endonucleases. It was the first of a new class of genetic polymorphism that is likely to be highly useful in chromosome mapping by linkage studies in families (see later).

Unfortunately, because of lack of tight linkage or adequate frequency of heterozygosity at the marker locus, the efficiency of diagnosis by the linkage approach is not as high as one would like for maximal usefulness. However, as the map becomes "saturated," and especially as a large number of highly polymorphic nucleic acid polymorphisms (see later) become available, the usefulness of the approach should be great. Potentially, premorbid diagnosis of a disorder such as Huntington's disease is possible by the same approach as that used for prenatal diagnosis.

Future Progress and Application. It is likely that mapping of the human genome will progress very rapidly during the rest of this vintade. Indeed, Ruddle [2] suggests that the human genome may be mapped in complete chemical detail by the end of the century. Sequencing the nucleic acids of the gene has become, in many instances, easier than sequencing the amino acids of the gene product.

Family linkage studies of classic type are likely to have a renaissance through use of nucleic acid polymorphisms as marker traits [17]. "Restriction fragment length polymorphisms" (RFLPs), resulting from substitution of nucleic acids in either coding or noncoding segments, thereby changing the susceptibility to digestion by various restriction enzymes, are, it seems, frequent in the human genome. The markers can be mapped to specific chromosomes by somatic cell hybridization. The technology already realized, as well as potential, should permit testing for RFLPs in the DNA of cells contained in a small sample of blood. Family linkage studies using the RFLPs as marker traits will allow mapping of rare dominants of great medical interests, such as Huntington's disease, achondroplasia and the Marfan syndrome, for which the basic biochemical defect is not yet known and for which there is also no specific marker at the cellular level which can be used for mapping by somatic cell hybridization. The gene pathology for some of these disorders may well be defined before the defective gene product is identified. The usefulness in premorbid diagnosis, identification of heterozygous carrier of recessive disorders and prenatal diagnosis is obvious. The role of specific genetic variation in determining susceptibility to common disorders, such as hypertension, diabetes and congenital malformations, is likely to be clarified. Detailed knowledge of the anatomy of the human genome is likely to elucidate normal regulatory mechanisms involved in differentiation and all aspects of genetic function. Mapping information and knowledge of regulatory mechanisms will probably be critically important to efforts to replace mutant genes with normal ones. Capability of complete genotyping of any person is a more remote but no less realistic expectation.

REFERENCES

1. McKusick VA, Ruddle FH: The status of the gene map of the human chromosomes. Science 1977; 196: 390–405.
2. McKusick VA, Ruddle FH: The status of the gene map of the human chromosome—1980. Science (in press).
3. Siniscalco M: Approaches to human linkage. In: Steinberg, AG, Bearn AG, Motulsky AG, Childs B, eds.: Progr Med Genet 1979; 3(NS): 221–307.
4. Conneally PM, Rivas ML: Linkage analysis in man. In: Harris H, Hirschhorn K, eds: Advances in Human Genetics, vol. 10, New York: Plenum, 1980.
5. McKusick VA: Mendelian inheritance in man. Catalogs of autosomal dominant, autosomal recessive and X-linked phenotypes, 5th ed. Baltimore: Johns Hopkins University Press, 1978.
6. Lundin LG: Evolutionary conservation of large chromosomal segments reflected in mammalian gene maps. Clin Genet 1979; 16: 72–81.
7. Edwards YH, Hopkinson DA: The genetic determination of fumarase isozymes in human tissues. Ann Hum Genet 1979; 42: 303–313.
8. Albring M, Griffith J, Attardi G: Association of a protein structure of probably membrane derivation with HeLa cell mitochondrial DNA near its origin of replication. Proc Natl Acad Sci 1977; 74: 1348–1352.
9. Beale G, Knowles J: Extranuclear genetics. Baltimore: University Park Press, 1978.
10. Crews ST, Ojala DK, Pasakovy JW, Nishiguchi J, Attardi G: Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. Nature 1979; 277: 192–198.
11. Fine PEM: Mitochondrial inheritance and disease. Lancet 1978; 2: 659–662.
12. Mitchell CH, Attardi G: Cytoplasmic transfer of chloramphenicol resistance in a human cell line. Somat Cell Genet 1978; 4: 734–744.
13. Harris H: Multilocus enzymes in man. In: Ciba Foundation Symposium No. 66 (NS): Human genetics: possibilities and realities. Amsterdam: Excerpta Medica, 1979.
14. Comi P, Giglioni B, Ottolenghi S, et al.: Globin synthesis in single erythroid bursts from cord blood: studies on gamma-to-beta and G-gamma-to-A-gamma switches. Proc Natl Acad Sci 1980; 77: 362–365.
15. Tonegawa S, Brack C, Hozumi N, Matthyssens G, Schuller R: Dynamics of immunoglobulin genes. Immunol Rev 1977; 36: 73–94.
16. Kan YW, Dozy AM: Antenatal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells. Lancet 1978; 2: 910–911.
17. Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 1980; 32: 314–331.
18. Forget BG: Molecular genetics of human hemoglobin synthesis. Ann Intern Med 1979; 91: 605–616.
19. Auerbach D, Rutter WJ, Shows TB: The insulin gene is located on chromosome 11 in human. Nature (in press).
20. Fritsch EF, Lawn RM, Maniatis T: Molecular cloning and characterization of the human B-like globin gene cluster. Cell 1980; 19: 959–972.
21. Evans JH, Hamerton JL, Klinger HP, McKusick VA (eds): Human gene mapping 5: Edinburgh Conference (1979), Fifth International Workshop on human gene mapping. Birth Defects 1979; 15: 1–236. Also Cytogenet Cell Genet 1979; 25: 1: 236.